

**In the United States Patent and Trademark Office**

Serial number: 09/728,689

Application Filed: Dec. 1, 2000

Applicants: Bax et. al.

Application Title: Technique for Extracting Data from Structured Documents

Examiner: Gregory J. Vaughn

Mailed: July 7, 2005

## Amendment A

Commissioner for Patents  
PO Box 1450  
Alexandria, VA 22313-1450

Sir:

## Amendments

In response to the Office Letter mailed January 31, 2005, and a June 6, 2005 phone conversation with Augusto Callejas, an iSpheres employee (323-841-6485), please amend the application as follows.

## SPECIFICATION

- Change the word “schematron” to “schemalyzer” where it refers to the invention. (The word “schematron” is in use to refer to a grammar-checker for XML, so iSpheres now uses the word “schemalyzer.”)

## CLAIMS

**Please replace the claims by the following claims:**

1. (CURRENTLY AMENDED) A method for extracting records from a structured text in a computer system, comprising:

identifying potential locations of values of record fields in the text by identifying locations in the text of items in lists of known potential values for record fields.

identifying a region of interest in the text by applying multiple candidate region partitioners, evaluating each to measure how well it isolates a region with a high density and a high amount of potential locations of values of record fields, selecting one that measures best, and applying it to produce a region of interest.

segmenting the region of interest into record regions that each contain data for a single record by applying multiple candidate segmenters, evaluating each to measure how well it segments into regions such that each region has one field value per record